

PAT-NO: JP02005278168A  
DOCUMENT-IDENTIFIER: JP 2005278168 A  
TITLE: METHOD AND SYSTEM FOR PROCESSING COMPRESSED  
INPUT VIDEO  
PUBN-DATE: October 6, 2005

INVENTOR-INFORMATION:

NAME	COUNTRY
XIN, JUN	N/A
ANTHONY, VETRO	N/A
SUN, HUIFANG	N/A

ASSIGNEE-INFORMATION:

NAME	COUNTRY
MITSUBISHI ELECTRIC RESEARCH LABORATORIES INC	N/A

APPL-NO: JP2005057390

APPL-DATE: March 2, 2005

PRIORITY-DATA: 2004791203 ( March 2, 2004)

INT-CL (IPC): H04N007/24, G09G005/00 , G09G005/391 , H04N007/01

ABSTRACT:

PROBLEM TO BE SOLVED: To provide a method and a system for processing a compressed input video.

SOLUTION: The compressed input video is processed to produce an interlaced picture and macroblock coding information of the input video. The interlaced picture has a first spatial resolution and a top-field and a bottom-field. The top-field and the bottom-field of the interlaced picture are filtered adaptively according to the macroblock coding information to produce a progressive picture with a second spatial resolution less than the

first  
spatial resolution.

COPYRIGHT: (C) 2006, JPO&NCIPI

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2005-278168

(P2005-278168A)

(43) 公開日 平成17年10月6日(2005.10.6)

(51) Int.Cl. <sup>7</sup>	F I	テーマコード (参考)
H04N 7/24	H04N 7/13	Z 5C059
G09G 5/00	H04N 7/01	Z 5C063
G09G 5/391	G09G 5/00	555A 5C082
H04N 7/01	G09G 5/00	520V

審査請求 未請求 請求項の数 20 O L 外国語出願 (全 27 頁)

(21) 出願番号	特願2005-57390 (P2005-57390)	(71) 出願人	597067574
(22) 出願日	平成17年3月2日(2005.3.2)		ミツビシ・エレクトリック・リサーチ・ラ
(31) 優先権主張番号	10/791203		ボラトリーズ・インコーポレイテッド
(32) 優先日	平成16年3月2日(2004.3.2)		アメリカ合衆国、マサチューセッツ州、ケ
(33) 優先権主張国	米国 (US)		ンブリッジ、ブロードウェイ 201
			201 BROADWAY, CAMBR
			IDGE, MASSACHUSETTS
			02139, U. S. A.
		(74) 代理人	100057874
			弁理士 曾我 道照
		(74) 代理人	100110423
			弁理士 曾我 道治
		(74) 代理人	100084010
			弁理士 古川 秀利

最終頁に続く

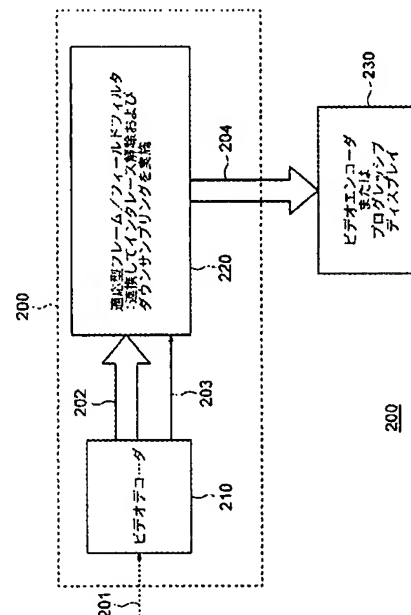
(54) 【発明の名称】 圧縮された入力ビデオを処理する方法及びシステム

## (57) 【要約】

【課題】 方法およびシステムは、圧縮された入力ビデオを処理する。

【解決手段】 圧縮された入力ビデオは処理されて、インタレースピクチャ、ならびに入力ビデオのマクロブロック符号化情報が生成される。インタレースピクチャは、第1空間解像度ならびにトップフィールドおよびボトムフィールドを有する。インタレースピクチャのトップフィールドとボトムフィールドは、マクロブロック符号化情報に従って適応的にフィルタリングされて、第1空間解像度より低い第2空間解像度を有するプログレッシブピクチャが生成される。

【選択図】 図2



## 【特許請求の範囲】

## 【請求項 1】

圧縮された入力ビデオを処理する方法であって、

前記圧縮された入力ビデオを復号し、第 1 空間解像度並びにトップフィールド及びボトムフィールドを有するインタレースピクチャ、並びに前記入力ビデオのマクロブロック符号化情報を生成することと、

前記インタレースピクチャのトップフィールド及びボトムフィールドを、前記マクロブロック符号化情報に従って適応的にフィルタリングし、前記第 1 空間解像度より低い第 2 空間解像度を有するプログレッシブピクチャを生成すること

を含む、圧縮された入力ビデオを処理する方法。

10

## 【請求項 2】

前記マクロブロック符号化情報は、マクロブロック符号化タイプ及びマクロブロック変換タイプを含む

請求項 1 記載の方法。

## 【請求項 3】

前記マクロブロック符号化タイプは、イントラ符号化及びインター符号化を含む

請求項 2 記載の方法。

## 【請求項 4】

前記マクロブロック変換タイプは、フレームベース変換及びフィールドベース変換を含む

20

請求項 2 記載の方法。

## 【請求項 5】

前記マクロブロック符号化情報は、前記マクロブロック符号化タイプがインター符号化であるときに、マクロブロック動きタイプ及び対応する動きベクトルをさらに含む

請求項 2 記載の方法。

## 【請求項 6】

前記マクロブロック動きタイプは、フレームベース動き及びフィールドベース動きを含む

請求項 5 記載の方法。

## 【請求項 7】

前記フィルタリングは、フレームベースフィルタリング及びフィールドベースフィルタリングを含む

30

請求項 1 記載の方法。

## 【請求項 8】

前記フィルタリングは、前記マクロブロック符号化タイプがインター符号化であり、かつ前記マクロブロック動きタイプがフィールドベースであるときには、フィールドベースである

請求項 7 記載の方法。

## 【請求項 9】

前記フィルタリングは、前記マクロブロック符号化タイプがインター符号化であり、前記マクロブロック動きタイプがフレームベースであり、かつ前記マクロブロックに対応する動きベクトルの絶対値がしきい値より大きいときには、フィールドベースである

40

請求項 7 記載の方法。

## 【請求項 10】

前記しきい値は、ゼロに等しい

請求項 9 記載の方法。

## 【請求項 11】

前記しきい値は、ゼロより大きい

請求項 9 記載の方法。

## 【請求項 12】

50

前記フィルタリングは、前記マクロブロック符号化タイプがイントラ符号化であり、かつ前記マクロブロック変換タイプがフィールドベースであるときには、フィールドベースである

請求項 7 記載の方法。

【請求項 13】

前記フィルタリングは、前記マクロブロック符号化タイプがイントラ符号化であり、かつ前記マクロブロック変換タイプがフレームベースであるときには、フレームベースである

請求項 7 記載の方法。

【請求項 14】

前記フィルタリングは、前記マクロブロック符号化タイプがインター符号化であり、前記マクロブロック動きタイプがフレームベースであり、かつ前記マクロブロックに対応する動きベクトルの絶対値がしきい値以下であるときには、フィールドベースである

請求項 7 記載の方法。

【請求項 15】

前記フィルタリングは、フレームベースであり、前記インタレースピクチャのトップフィールド及びボトムフィールドからの入力サンプルについて動作する

請求項 7 記載の方法。

【請求項 16】

前記フィルタリングは、フィールドベースであり、前記トップフィールド又はボトムフィールドからの入力サンプルについて動作する

請求項 7 記載の方法。

【請求項 17】

前記フィルタリングは、フィールドベースであり、前記ボトムフィールドからの入力サンプルについて動作する

請求項 7 記載の方法。

【請求項 18】

前記プログレッシブピクチャを出力ビデオに符号化することをさらに含む

請求項 1 記載の方法。

【請求項 19】

前記プログレッシブピクチャをディスプレイデバイス上にレンダリングすることをさらに含む

請求項 1 記載の方法。

【請求項 20】

圧縮された入力ビデオを処理するシステムであって、

前記圧縮された入力ビデオを復号し、第 1 空間解像度並びにトップフィールド及びボトムフィールドを有するインタレースピクチャ、並びに前記入力ビデオのマクロブロック符号化情報を生成する手段と、

前記インタレースピクチャのトップフィールド及びボトムフィールドを、前記マクロブロック符号化情報に従って適応的にフィルタリングし、前記第 1 空間解像度より低い第 2 空間解像度を有するプログレッシブピクチャを生成する手段と

を備える、圧縮された入力ビデオを処理するシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、包括的に、ビデオの処理に関し、より詳細には、インタレースビデオをインタレース解除するとともにダウンサンプリングすることに関する。

【背景技術】

【0002】

インタレースビデオでは、ビデオの各フレームは 2 つのフィールドを有する。一方のフ

10

20

30

40

50

フィールドは、フレームの全ての偶数番号のピクセルラインを含み、他方のフィールドは全ての奇数番号のピクセルラインを含む。2つのフィールドを一緒にインタレースすることによってビデオフレームが形成される。2つのフィールドは、動きの連続性をよりよくするために、インタレースディスプレイ上で交互に表示される。消費者TVセットの大多数はインタレースディスプレイデバイスである。インタレースビデオは、地上ビデオ放送、ケーブルテレビジョン(CATV)、ならびに、直接衛星放送(DBS)システムで広く使用される。現在のデジタルテレビジョン放送、特に高品位テレビジョン(HDTV)は主に、インタレースビデオを使用する。インタレースデジタルビデオの典型的な解像度は、比較的高く、たとえば、標準品位TV(SDTV)用で $720 \times 480$ 、HDTV用で $1920 \times 1080$ である。

10

## 【0003】

個人情報端末(PDA)および携帯電話を含む可搬型端末、ならびに、コンピュータモニターは通常、プログレッシブディスプレイを使用する。プログレッシブディスプレイでは、ビデオフレームの全てのピクセルラインは、上部から下部へ順次表示される。さらに、多くのプログレッシブディスプレイ、特にPDAと携帯電話は、表示能力が制限され、たとえば、 $320 \times 240$ が、現在、PDA用のハイエンドディスプレイであり、携帯電話のディスプレイ解像度は、一般に、さらに低い。

## 【0004】

MPEG-2は、放送業界によって現在使用されているビデオ符号化規格である。この規格は、インタレース式とプログレッシブの両方の高解像度デジタルビデオを効率よく表現することが可能である。

20

## 【0005】

MPEG-2ビデオは通常、2つのフィールドが一緒に符号化される「フレームピクチャ」を使用して符号化される。MPEG-2構文は、フィールドがフィールドピクチャとして個々に符号化される「フィールドピクチャ」の符号化もサポートする。本発明では、以下の説明でMPEG-2フレームピクチャを使用するが、説明はフィールドピクチャにも当てはまる。

## 【0006】

MPEG-2ビデオ符号化プロセスは、YCbCr色空間で表現されるビデオフレームについて動作する。画像が24ビットRGB形式で記憶される場合、画像は、最初に、YCbCr形式に変換されなければならない。各ビデオフレームは、重ならないマクロブロックに分割される。各マクロブロックは $16 \times 16$ ピクセルをカバーする。各マクロブロックは、4つの $8 \times 8$ 輝度(luma)(Y)ブロック、および対応する2つの $8 \times 8$ 色(chroma)ブロック(1つはCbブロックで、1つはCrブロック)を含む。マクロブロックは、動き補償予測(MCP)用の基本単位であり、ブロックは、離散コサイン変換(DCT)を適用するための基本単位である。

30

## 【0007】

MPEG-2ビデオには3つのフレームタイプ、すなわち、イントラフレーム(Iフレーム)、予測フレーム(Pフレーム)、および双方向予測フレーム(Bフレーム)がある。Iフレームは、他のフレームを参照せずに独立して符号化される。Iフレームのマクロブロックは、フレームDCTか、フィールドDCTのいずれかを使用することができる。Pフレームは、前の参照フレームに対して符号化される。マクロブロックを、イントラマクロブロックまたはインターマクロブロックとして符号化することができる。イントラマクロブロックは、Iフレーム内のマクロブロックのように符号化される。

40

## 【0008】

インターマクロブロックを、フレーム予測か、または、フィールド予測することができる。フレーム予測では、マクロブロックは、動きベクトルによって配置された基準フレーム内のブロックから予測される。フィールド予測では、マクロブロックは、2つの $16 \times 8$ ブロックに分割され、一方のブロックはトップフィールドに属し、他方のブロックはボトムフィールドに属する。各 $16 \times 8$ ブロックは、基準フレームのトップか、ボトムのか

50

ずれのフィールドが予測として使用されるかを指定するフィールド選択ビット、および、適切なフィールド内の16×8ピクセル領域を指す動きベクトルを有する。マクロブロックは、マクロブロックがゼロの動きベクトルおよび全てゼロの誤差項(error terms: 数値、条件)を有する時に省略されることができる。

#### 【0009】

Bフレームは、前の基準フレームと将来の基準フレームの両方に対して符号化される。Bフレームの符号化は、動きベクトルが将来の基準フレーム内のエリアを参照することができることを除いて、Pフレームと同様である。

#### 【0010】

通常、プログレッシブ可搬型デバイス上で表示するために、MPEG-2符号化ビデオは、MPEG-4シンプルプロファイル(SP)などの低解像度プログレッシブビデオに最適化された形式に符号変換(transcode)される必要がある。

#### 【0011】

MPEG-2符号化インタレースビデオが、MPEG-4 SPのような低解像度プログレッシブビデオに符号変換される時か、または、低解像度プログレッシブディスプレイ上に表示される時に2つの問題が発生する。1つの問題は、エイリアシング、鋸歯タイプエッジ、およびラインフリッカを含む、よく知られているインタレース時のアーチファクトによるものである。もう1つの問題は、解像度のミスマッチによるものである。インタレース解除およびダウンサンプリングフィルタリングは、2つの問題を解決するための従来の技法である。

#### 【0012】

基本的なインタレース解除方法は、米国特許第4,750,057号、第4,800,436号、第4,881,125号、第5,748,250号、および第6,661,464号に記載されているような「weave」、「bob」、「discard」、および「adaptive」を含む。「weave」法は、フレームの2つのフィールドを一緒にインタレースするだけである。処理されたビデオは、インタレースアーチファクトを有するが、フル解像度を有する。「bob」法は、全てのフィールドを個々のフレームとして表示する。したがって、フレームレートは2倍であるが、全てのフレームで、空間解像度が失われる。「discard」法は、1つおきにフィールドを廃棄し、したがって、インタレースアーチファクトは完全になくなるが、解像度の半分が失われ、動きが流動的に見えない。「adaptive」法は、「weave」法と「bob」法を組み合わせる。「adaptive」法は、インタレースアーチファクトが存在する時のみにインタレース解除を実施し、その他の場合には、「weave」法を使用する。

#### 【0013】

通常、インタレースアーチファクトは、動きを有する領域のみがインタレース解除を必要とするため、動き情報を使用して検出される。「adaptive」法は、「weave」または「bob」より優れた性能を達成することができるが、動き検出は、通常、計算に関し高価であり、システムコストをかなり増加させる。動き補償式のインタレース解除法などの進歩した方法は、計算上の複雑さをますます大きくしながら、よりよい品質を達成することができる。米国特許第5,784,115号および第6,442,203号を参照されたい。

#### 【0014】

解像度のミスマッチに対処するために、ダウンサンプリングが実施される必要がある。一般的な連結した補間-間引き、ならびに、他のもっと進歩した方法を、この目的のために適用することができる。米国特許第5,289,292号、第5,335,295号、第5,574,572号、第6,175,659号、および第6,563,964号を参照されたい。

#### 【0015】

図1は、従来技術のシステム100の1つの例を示す。ビデオデコーダ110は、圧縮されたインタレースビデオ101を復号し、復号されたインタレースピクチャ102をイ

インタレース解除器120に送る。インタレース解除されたプログレッシブピクチャ103は、ダウンサンプリングフィルタ130によってダウンサンプリングされる。最後に、インタレース解除され、ダウンサンプリングされたピクチャ104は、エンコーダ140、プログレッシブディスプレイデバイス、または他の処理に渡される。ダウンサンプリング130はフル解像度のインタレース解除されたピクチャに対して実施されるため、不必要な付加的計算が導入される可能性がある。

【発明の開示】

【発明が解決しようとする課題】

【0016】

その結果、高解像度のインタレースコンテンツを低解像度プログレッシブディスプレイ上に表示するために、インタレース解除とダウンサンプリングを連携して実施することに対する必要性が存在する。計算上の複雑さがかなり低く、ビデオ品質を費用効果的に改善することができる、MPEG-2インタレース解除およびダウンサンプリングシステムに対する必要性も存在する。

【課題を解決するための手段】

【0017】

本発明は、圧縮された入力ビデオの処理を行う。圧縮された入力ビデオは処理されて、インタレースピクチャ、ならびに入力ビデオのマクロブロック符号化情報が生成される。インタレースピクチャは、第1空間解像度ならびにトップフィールドおよびボトムフィールドを有する。インタレースピクチャのトップフィールドとボトムフィールドは、マクロブロック符号化情報に従って適応的にフィルタリングされて、第1空間解像度より低い第2空間解像度を有するプログレッシブピクチャが生成される。

【発明を実施するための最良の形態】

【0018】

本発明は、表示、再符号化、または他の処理のために、復元されたビデオを連携してインタレース解除するとともにダウンサンプリングするシステムおよび方法を提供する。本発明では、適応型フレーム/フィールドフィルタリングプロセスを使用して、インタレース解除とダウンサンプリングを連携して実施する。本発明は、入力圧縮ビデオが、MPEG-2フレームピクチャで符号化される時に特に有用であり、MPEG-2フレームピクチャは、現在、ビデオを放送し記憶するための、有力なビデオ符号化法である。しかしながら、本発明が、MPEG-2フレームピクチャ以外の他のビデオ符号化システムにも適用されることが理解されるべきである。

【0019】

システム構造

図2は、本発明に従って、インタレース解除とダウンサンプリングを連携して実施する、適応型フィルタリングシステム200を示す。ビデオデコーダ210は、MPEG-2符号化インタレースビデオ201を復号し、インタレースビデオピクチャ202、ならびに、マクロブロック符号化モード、符号化タイプ、および動き情報を含む、付随(side)情報203を再構成する。

【0020】

適応型フィルタ

適応型フィルタ220は、インタレースアーチファクトを検出するために付随情報203を使用する。付随情報はマクロブロックと関連するため、インタレースアーチファクト検出および適応型フィルタリング220もまた、マクロブロックベースで適用される。適応型フレーム/フィールドフィルタリングは、インタレースアーチファクトに応じて適用される。

【0021】

マクロブロックベースの処理は、フィルタリング結果204を、マクロブロックが処理された直後に出力することができるため、遅延が少ないという付加的な利点を有する。インタレース解除され、ダウンサンプリングされたピクチャ204は、その後、さらなる処



理230、たとえば、再符号化およびプログレッシブディスプレイ上での表示のために送り出される。

#### 【0022】

フレームフィルタリングは、フレームからのサンプルを使用し、フィールドフィルタリングは、1つのフィールドからのサンプルのみを使用する。フレームフィルタリングは、インタレースアーチファクトが存在しないピクセル領域で使用され、フィールドフィルタリングは、インタレースアーチファクトが実際に存在するピクセル領域で使用される。圧縮された入力ビデオ201から復号された、符号化モードおよび/または動きベクトルを含む付随情報に基づいて、ピクセル領域についてのインタレースアーチファクトの目安が確定され、適応型フレーム/フィールドフィルタリングが、ピクセル領域に応じて適用される。

10

#### 【0023】

圧縮MPEG-2ビデオストリームで符号化された多くの符号化情報(decisions)は、インタレースアーチファクトの存在を検出し、フレームフィルタリングか、フィールドフィルタリングのいずれを適用すべきかについて決定を行うのに役立つ可能性がある。

#### 【0024】

##### フィルタ決定方法

図3は、復号された付随情報203を使用して、適切なフィルタリングの決定を行う方法220を詳細に示す。特に有用な付随情報は、マクロブロック符号化タイプおよびマクロブロック変換タイプを指示するマクロブロックレベルの符号化パラメータを含む。マクロブロック符号化タイプは、特定の符号化形式によって規定される、時間的に予測し符号化する全てのモードを含む、「イントラ」か、「インター」のいずれかであることができ、一方、変換タイプは、変換が、マクロブロックの、フレームブロック構成またはフィールドブロック構成について動作することを意味する、フレームベースか、フィールドベースのいずれかであることができる。マクロブロックがインター符号化される場合、マクロブロック動きタイプもまた、フィルタ決定中に考慮される。この決定のために、動きタイプは、動き補償予測が、フレームベースか、または、フィールドベースかを指示する。

20

#### 【0025】

入力ビデオが、MPEG-2形式で符号化される実施形態では、MACROBLOCKTYPEは符号化タイプを示し、DCTTYPEは変換タイプを示し、MOTIONTYPEは動きタイプを示す。他のビデオ符号化形式の対応する構文要素について、同様なマッピングが示されるであろう。

30

#### 【0026】

イントラ符号化されたマクロブロックの場合、フィールドベース変換、たとえば、MPEG-2構文でDCTTYPE=1の使用は、インタレースアーチファクトが存在することを指示する傾向がある。イントラ符号化されたマクロブロックの場合、動きが存在する時には、通常、インタレースアーチファクトが存在する。したがって、フィールドベース予測、たとえば、MPEG-2構文でMOTIONTYPE=「フィールドベース予測」は、インタレースアーチファクトが存在することを意味する可能性が強く、一方、フレームベース予測、たとえば、MPEG-2構文でMOTIONTYPE=「フレームベース予測」は同様に、マクロブロックがゼロの動きベクトルか、または、非常に小さい動きベクトルを持っていないければ、インタレースアーチファクトを指示する。

40

#### 【0027】

インタレースアーチファクト検出および適応型フィルタリング法220は、ピクチャの全ての入力マクロブロック301に適用されて、出力マクロブロック302が生成される。方法220への入力301は、実際のビデオデータ202および各マクロブロックの付随情報203からなることに留意されたい。

#### 【0028】

方法220での最初の決定は、符号化タイプを確定すること310である。タイプがインター符号化である場合、動きタイプを確定する320。インター符号化マクロブロック

50

について、動きタイプがフィールドベースである場合、フィールドベースフィルタリングを適用して360、出力マクロブロック302が生成される。しかしながら、動きタイプがフレームベースである場合、動きベクトル(VM)の大きさがしきい値(T)より大きいかが判断される330。ここで、しきい値はゼロか、または、非ゼロ値にセットされ得る。マクロブロックがインター符号化され、動きタイプがフレームベースである場合、MVがT以下であれば、フィールドベースフィルタリングが適用され360、MVがTより大きい時、フレームベースフィルタリングが適用される370。

#### 【0029】

一方、符号化タイプ310がイントラ符号化である場合、変換タイプが確定される340。イントラ符号化マクロブロックについて、変換タイプ340がフィールドベースである場合、フィールドベースフィルタリングを適用して360、出力マクロブロック302が生成される。しかしながら、変換タイプ340がフレームベースである場合、フレームベースフィルタリングを適用して370、出力マクロブロック302が生成される。

#### 【0030】

##### フレーム／フィールドフィルタリング

図4は、フィルタリングの前と後での、部分的なマクロブロック400の相対的なサンプル位置の一例を示す。ここで、水平と垂直の両方の次元で、2のダウンサンプリング比を仮定する。図4では、記号は、輝度入力／トップフィールド401、輝度入力／ボトムフィールド402、色入力／トップフィールド403、色入力／ボトムフィールド404、輝度出力405、および色出力406である。フレームベースまたはフィールドベースフィルタリングは、より低い次元のサンプリンググリッドでサンプル出力を生成する。出力サンプルの位置は、入力ピクセル値を処理するのに使用されるフィルタの係数によって効率よく決まる。出力される輝度および色サンプルの相対的な位置が維持される、すなわち、出力サンプリンググリッドの構造が、解像度は低い、入力サンプリンググリッドの構造と同じであるように、フィルタリングを実施することが望ましい。

#### 【0031】

図5は、水平と垂直の両方の次元での2のダウンサンプリング比について、この出力配置を達成することができる、フレームベースおよびフィールドベースフィルタリング動作の例を示す。図5では、マクロブロックの部分500が示される。記号A、B、C、D、E、E、G、H501はトップフィールドの入力サンプルを示し、記号a、b、c、d、e、f、g、h502はボトムフィールドの入力サンプルを示し、記号 $\alpha$ 、 $\beta$ 、 $\gamma$ 、 $\delta$ 503は出力サンプルである。双1次補間を使用して、フレームフィルタリングは $\alpha = (A + B + a + b) / 4$ に従って実施され、フィールドフィルタリングは $\alpha = (3 \times (A + B) + E + F) / 8$ に従って実施される。同様に、入力サンプルから、他の出力サンプルを計算することができる。

#### 【0032】

水平および／または垂直の次元で2以外のダウンサンプリング比を有する場合、適応型フィルタ係数および出力サンプル位置は、出力サンプリンググリッド構造が、解像度は低い、入力サンプリンググリッド構造と同じであるように確定される。2分の1のダウンサンプリングの場合と同様に、双1次補間を使用して、入力サンプルに適用されるフィルタ係数、すなわち、重み係数が確定される。同様に出力サンプルを所望のサンプル位置に配置することを可能にする、改善された周波数応答を有する、より高性能のフィルタを使用してもよい。

#### 【0033】

本発明の精神および範囲内で、種々の他の適応および変更を行うことができることが理解されるべきである。したがって、本発明の真の精神および範囲内に入る全ての変形および変更を包含することが、添付特許請求項の目的である。

#### 【図面の簡単な説明】

#### 【0034】

【図1】従来技術のインタレース解除器およびダウンサンプリングフィルタのブロック図

である。

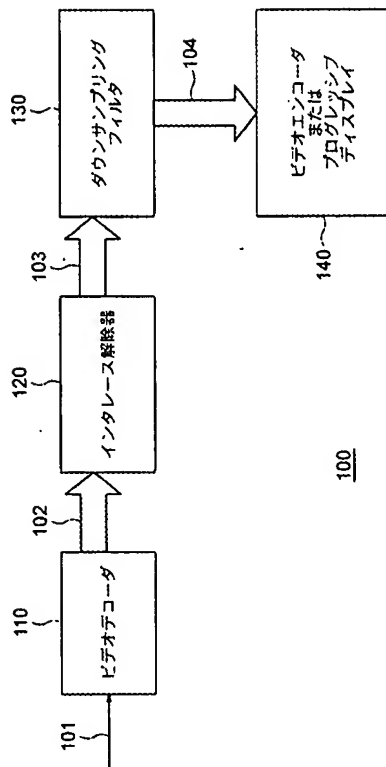
【図 2】 本発明による適応型フィルタリングシステムのブロック図である。

【図 3】 本発明による適応型フレームフィルタリング方法のフロー図である。

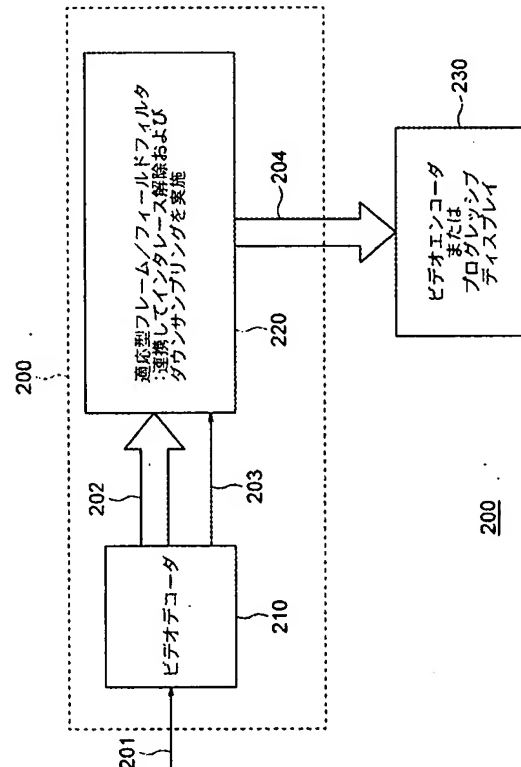
【図 4】 本発明による適応型フィルタリングの前と後での、輝度および色サンプルの相対位置のブロック図である。

【図 5】 図 4 の出力サンプルを生成するための、フレームフィルタリングおよびフィールドフィルタリングのブロック図である。

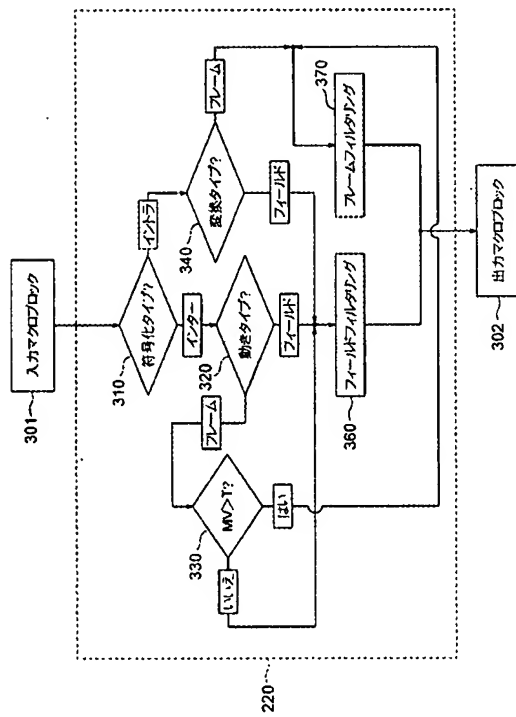
【図 1】



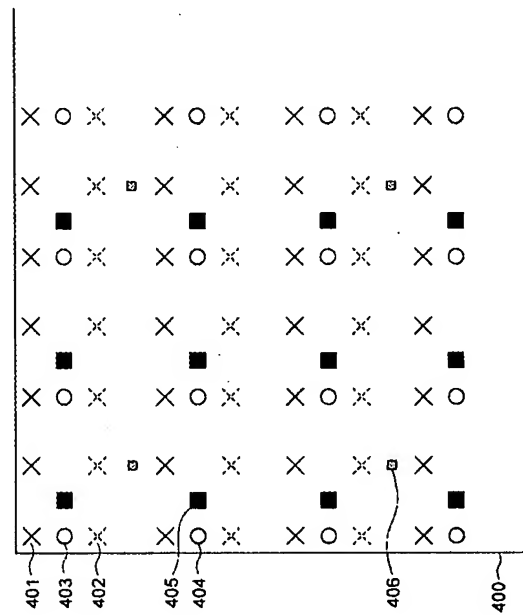
【図 2】



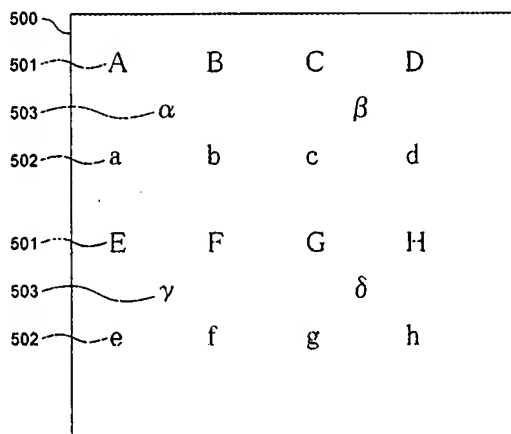
【図 3】



【図 4】



【図 5】



## フロントページの続き

(74)代理人 100094695

弁理士 鈴木 憲七

(74)代理人 100111648

弁理士 梶並 順

(72)発明者 ジュン・シン

アメリカ合衆国、マサチューセッツ州、クインシー、サザン・アーテリー 1071、アパートメント 204

(72)発明者 アンソニー・ヴェトロ

アメリカ合衆国、マサチューセッツ州、ケンブリッジ、ハーバード・ストリート 353、ユニット 45

(72)発明者 ハイファン・スン

アメリカ合衆国、マサチューセッツ州、ビレライカ、スカイライン・ドライブ 2

Fターム(参考) 5C059 LA05 LB05 MA00 MA23 NN21 TA13 TC12 TC27 TD12 UA05

UA11

5C063 BA04 CA07 CA40

5C082 AA02 BA12 BA29 BA34 BA35 BA42 BB44 BC19 CA34 CA84

DA26 MM05

【外国語明細書】

## **Title of the Invention**

Method and System for Processing Compressed Input Video

## **Detailed Description of Invention**

### **Field of the Invention**

The invention relates generally to video processing, and more particularly to de-interlacing and downsampling of interlaced video.

### **Background of the Invention**

In an interlaced video, each frame of the video has two fields. One field includes all even pixel lines of the frame, and the other frame includes all odd pixel lines. Interlacing the two fields together forms the video frame. The two fields are displayed alternatively on an interlaced display for better motion continuity. The majority of consumer TV sets are interlaced display devices. Interlaced video is widely used in terrestrial video broadcasting, cable television (CATV) as well as direct broadcast satellite (DBS) systems. The current digital television broadcasting, and particularly high definition television (HDTV) mainly uses interlaced video. Typical resolutions of digital interlaced video are relatively high, e.g., 720 x 480 for standard definition TV (SDTV) and 1920 x 1080 for HDTV.

Portable terminals, including personal digital assistants (PDA) and cell phones, and computer monitors typically use progressive display. In progressive display, all pixel lines of a video frame are displayed sequentially from top to bottom. In

addition, many progressive displays, PDA and cell phones in particular, have limited display capability, e.g., 320 x 240 is currently the high-end display for PDA, and cell phone display resolution generally is even smaller.

MPEG-2 is a video coding standard currently used by broadcasting industry. This standard is capable of efficiently representing high-resolution digital video, both interlaced and progressive.

MPEG-2 video is usually encoded using 'frame pictures', where the two fields are coded together. The MPEG-2 syntax also supports coding of 'field-pictures' where the fields are coded separately as field pictures. We use MPEG-2 frame-picture in the following descriptions, but the description also applies to field-picture.

The MPEG-2 video-coding process operates on video frames represented in the YCbCr color space. If images are stored in a 24-bit RGB format, then the images must first be converted to the YCbCr format. Each video frame is divided into non-overlapping macroblocks. Each macroblock covers a 16 x 16 pixels. Each macroblock includes four 8 x 8 luma (Y) blocks, and two corresponding 8 x 8 chroma blocks (one Cb block and one Cr block). Macroblocks are the basic units for motion compensated prediction (MCP), and blocks are the basic units for applying discrete cosine transform (DCT).

There are three types of frames in the MPEG-2 video: intra-frames (I-frames), predicted frames (P-frames), and bi-directional predicted frames (B-frames). An I-frame is coded independently without referring to other frames. A macroblock in an I-frame can use either frame-DCT or field-DCT. A P-frame is coded relative to

a prior reference frame. A macroblock can be coded as an intra-macroblock or an inter-macroblock. An intra-macroblock is encoded like a macroblock in an I-frame.

An inter-macroblock can be frame-predicted or field-predicted. In frame-prediction, the macroblock is predicted from a block in the reference frame positioned by a motion vector. In field-prediction, the macroblock is divided into two 16 x 8 blocks, one block belongs to the top field, and the other block belongs to the bottom field. Each 16 x 8 block has a field selection bit, which specifies whether the top or the bottom field of the reference frame is used as prediction, and a motion vector, which points to the 16 x 8 pixel region in the appropriate field. A macroblock can be skipped when it has a zero motion vector and all-zero error terms.

A B-frame is coded relative to both a prior reference frame and a future reference frame. The encoding of a B-frame is similar to a P-frame, except that the motion vectors can refer to areas in the future reference frame.

Typically, for display on progressive portable devices, MPEG-2 coded video needs to be transcoded to a format optimized for low-resolution progressive video such as MPEG-4 simple profile (SP).

Two problems arise when MPEG-2 coded interlaced video is transcoded to a low-resolution progressive video like MPEG-4 SP, or when it is to be displayed on low-resolution progressive display. One problem is due to well-known interlacing artifacts, including aliasing, saw-tooth type edge-distortion and line flicker. The other problem is due to a resolution mismatch. De-interlacing and downsampling filtering are conventional techniques to solve the two problems.



Basic de-interlacing methods include "weave," "bob," "discard," and "adaptive," as in U.S. Patent No. 4,750,057, 4,800,436, 4,881,125, 5,748,250, and 6,661,464. The "weave" method only interlaces the two fields of a frame together. The processed video has interlacing artifacts but with full resolution. The "bob" method displays every field as individual frames. Thus, the frame rate doubles, but the spatial resolution is lost in every frame. The "discard" method discards every other field, and therefore the interlacing artifacts are completely eliminated, but half of the resolution is lost and motion does not appear as fluid. The "adaptive" method combines the "weave" and "bob" methods. It performs de-interlacing only when there are interlacing artifacts, and uses the "weave" method elsewhere.

Typically, the interlacing artifacts are detected using motion information because only regions with motion need de-interlacing. Although the "adaptive" method can achieve better performance than "weave" or "bob," the motion detection is usually computationally expensive and significantly increases the system cost. Advanced methods such as motion compensated de-interlacing methods can achieve better quality with even greater computational complexity, see U.S. Patent No. 5,784,115, and 6,442,203.

To deal with the resolution mismatch, downsampling needs to be performed. Generic concatenated interpolating-decimating, as well as other more advanced methods, can be applied for this purpose, see U.S. Patent No. 5,289,292, 5,335,295, 5,574,572, 6,175,659, and 6,563,964.

Figure 1 shows one example prior art system 100. A video decoder 110 decodes a compressed interlaced video 101 and sends decoded interlaced pictures 102 to a de-interlacer 120. De-interlaced progressive pictures 103 are downsampled 130 by

a downsampling filter. Finally, the de-interlaced and downsampled pictures 104 are passed on to an encoder 140, progressive display device, or other processing. Because the downsampling 130 is performed on the full-resolution de-interlaced pictures, unnecessary additional computations can be introduced.

Consequently, there exists a need for jointly performing de-interlacing and downsampling for displaying high-resolution interlaced content on low-resolution progressive display. There is also a need for an MPEG-2 de-interlacing and downsampling system that has a comparatively low computational complexity and can improve video quality cost effectively.

### **Summary of the Invention**

The invention provides for processing a compressed input video. The compressed input video is processed to produce an interlaced picture, and macroblock coding information of the input video. The interlaced picture has a first spatial resolution, and a top-field and a bottom-field. The top-field and the bottom-field of the interlaced picture are filtered adaptively according to the macroblock coding information to produce a progressive picture with a second spatial resolution less than the first spatial resolution.

### **Detailed Description of the Preferred Embodiment**

Our invention provides a system and method for jointly de-interlacing and downsampling decompressed video for display, re-encoding or other processing. We perform the de-interlacing and downsampling jointly using an adaptive frame/field filtering process. Our invention is particularly useful when the input

compressed video is coded in MPEG-2 frame-pictures, which is currently the dominant video coding method for broadcasting and storage video. However, it should be understood that this invention also applies to other video coding systems other than MPEG-2 frame-pictures.

### **System Structure**

Figure 2 shows an adaptive filtering system 200 that jointly performs de-interlacing and downsampling according to our invention. A video decoder 210 decodes an MPEG-2 coded interlaced video 201 and reconstructs interlaced video pictures 202 as well as side information 203, including macroblock coding-modes, coding-types and motion information.

### **Adaptive Filter**

An adaptive filter 220 uses the side information 203 for detecting interlacing artifacts. Because the side information is associated with macroblocks, the interlacing artifacts detection and adaptive filtering 220 are also applied on a macroblock basis. Adaptive frame/field filtering is applied according to the interlacing artifacts.

Macroblock-based processing has the additional advantage of low-delay, because the filtering result 204 can be outputted immediately after a macroblock is processed. The de-interlaced and downsampled pictures 204 are then sent out for further processing 230, for example, re-encoding and displaying on a progressive display.

Frame filtering uses samples from the frame, and field filtering uses only samples from one field. Frame filtering is used at pixel regions where no interlacing artifacts are present and field filtering is used at pixel regions where interlacing artifacts do exist. Based on the side information decoded from the compressed input video 201, including coding modes and/or motion vectors, indications of interlacing artifacts for a pixel region are determined, and adaptive frame/field filtering is applied accordingly to the pixel region.

Many coding decisions encoded in the compressed MPEG-2 video stream can be useful for detecting the existence of interlacing artifacts and to make the decision whether to apply frame filtering or field filtering.

### **Filter Decision Method**

Figure 3 shows in detail the method 220 that uses the decoded side information 203 to make the appropriate filtering decisions. Particularly useful side information includes the macroblock-level coding-parameters that indicate the macroblock coding type and the macroblock transform type. The macroblock coding type can be either 'intra' or 'inter' including all temporally predictive-coding modes defined by the particular coding format, while the transform type can be either frame-based or field-based meaning that the transform operates on frame-block or field-block configurations of a macroblock. If the macroblock is inter-coded, a macroblock motion type is also considered during the filter decision. For the purpose of this decision, the motion type indicates whether the motion compensated prediction is frame-based or field-based.

In an embodiment in which the input video is coded in an MPEG-2 format, the `MACROBLOCK_TYPE` indicates the coding type, the `DCT_TYPE` indicates the transform type, and the `MOTION_TYPE` indicates the motion type. A similar mapping could be shown for the corresponding syntax elements in other video coding formats.

For intra-coded macroblocks, the use of a field-based transform, e.g., `DCT_TYPE` = 1 in MPEG-2 syntax, tends to indicate interlacing artifacts exist. For intra-coded macroblocks, there are typically interlacing artifacts when there is motion. Therefore, field-based prediction, e.g., `MOTION_TYPE` = "Field-based-prediction" in the MPEG-2 syntax, very likely means there are interlacing artifacts, while frame-based prediction, e.g., `MOTION_TYPE` = "Frame-based-prediction" in the MPEG-2 syntax also indicates interlacing artifacts unless the macroblock has zero motion-vector or very small motion vector.

The interlacing artifacts detection and adaptive filtering method 220 is applied to all input macroblocks 301 in a picture to produce output macroblocks 302. Note that the input 301 to the method 220 is comprised of actual video data 202 and side information 203 of each macroblock.

The first decision in the method 220 is to determine 310 the coding type. If the type is inter-coded, determine 320 the motion type. For inter-coded macroblocks, if the motion type is field-based, then apply field-based filtering 360 to produce the output macroblock 302. However, if the motion-type is frame-based, then determine 330 whether the magnitude of motion vectors (MV) is greater than a threshold (T), where the threshold may be set to zero or a non-zero value. For the case when the macroblock is inter-coded and the motion type is frame based, we

apply field-based filtering 360 if MV is less than or equal to T, and frame-based filtering 370 when MV is greater than T.

On the other hand, if the coding type 310 is intra-coded, determine 340 the transform type. For intra-coded macroblocks, if the transform type 340 is field-based, then apply field-based filtering 360 to produce output macroblock 302. However, if the transform type 340 is frame-based, then apply frame-based filtering 370 to produce the output macroblock 302.

### **Frame/Field Filtering**

Figure 4 shows one example of relative sample-positions of a partial macroblock 400 before and after filtering, where a down-sampling ratio of 2 in both the horizontal and vertical dimensions is assumed. In Figure 4, the symbols are luma-input/top-field 401, luma-input/bottom-field 402, chroma-input/top-field 403, chroma-input/bottom-field 404, the luma-output 405, and the chroma-output 406. The frame-based or field-based filtering produces output samples in a lower-dimension sampling grid. The positions of the output samples effectively depend on the filter coefficients that are used to process the input pixel values. It is desirable to perform the filtering so that the relative positions of output luma and chroma samples are maintained, i.e., the structure of the output sampling grid is the same as the structure of the input sampling grid, but with less resolution.

Figure 5 shows examples of frame-based and field-based filtering operations that can achieve this output positioning for a down-sampling ratio of 2 in both the horizontal and vertical dimensions. In Figure 5, a portion 500 of a macroblock is shown. Symbols A, B, C, D, E, F, G, H 501 indicates input samples of the top field,

symbols a, b, c, d, e, f, g, h 502 are input samples of the bottom field, symbols  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$  503 are output samples. Then, using bilinear interpolation, the frame filtering is performed according to  $\alpha = (A+B+a+b)/4$ , and field filtering according to  $\alpha = (3 \times (A+B) + E + F)/8$ . Likewise, other output samples can be computed from input samples.

For the case with down-sampling ratio other than 2 in horizontal and/or vertical dimensions, adaptive filter coefficients and output sample positions are determined so that the structure of the output sampling grid is the same as the structure of the input sampling grid, but with less resolution. As with the down-sampling by a factor of 2, bilinear interpolation is used to determine the filter coefficients, i.e., weighting factors, that are applied to input samples. More sophisticated filters with improved frequency response that also provide output samples at the desired sample positions may also be used.

It is to be understood that various other adaptations and modifications can be made within the spirit and scope of the invention. Therefore, it is the object of the appended claims to cover all such variations and modifications as come within the true spirit and scope of the invention.

**Brief Description of the Drawings**

Figure 1 is a block diagram of a prior art de-interlacer and downsampling filter;

Figure 2 is a block diagram of an adaptive filtering system according to the invention;

Figure 3 is a flow diagram of an adaptive frame filtering method according to the invention;

Figure 4 is a block diagram of relative positions of luma and chroma samples before and after adaptive filtering according to the invention; and

Figure 5 is a block diagram frame filtering and field filtering to produce the output samples of Figure 4.



1. A method for processing a compressed input video, comprising:
  - decoding the compressed input video to produce an interlaced picture, and macroblock coding information of the input video, the interlaced picture having a first spatial resolution, and a top-field and a bottom-field; and
  - filtering adaptively the top-field and the bottom-field of the interlaced picture according to the macroblock coding information to produce a progressive picture with a second spatial resolution less than the first spatial resolution.
2. The method of claim 1, in which the macroblock coding information includes a macroblock coding type and a macroblock transform type.
3. The method of claim 2, in which the macroblock coding type includes intra-coding and inter-coding.
4. The method of claim 2, in which the macroblock transform type includes a frame-based transform and a field-based transform.
5. The method of claim 2, in which the macroblock coding information further includes a macroblock motion type and corresponding motion vector when the macroblock coding type is inter-coding.
6. The method of claim 5, in which the macroblock motion type includes frame-based and field-based.

7. The method of claim 1, in which the filtering includes frame-based filtering and field-based filtering.
8. The method of claim 7, in which the filtering is field-based when the macroblock coding type is inter-coding and the macroblock motion type is field-based.
9. The method of claim 7, in which the filtering is field-based when the macroblock coding type is inter-coding, the macroblock motion type is frame-based, and the absolute value of motion vectors corresponding to the macroblock are greater than a threshold.
10. The method of claim 9, in which the threshold equals zero.
11. The method of claim 9, in which the threshold is greater than zero.
12. The method of claim 7, in which the filtering is field-based when the macroblock coding type is intra-coding and the macroblock transform type is field-based.
13. The method of claim 7, in which the filtering is frame-based when the macroblock coding type is intra-coding and the macroblock transform type is frame-based.
14. The method of claim 7, in which the filtering is frame-based when the macroblock coding type is inter-coding and the macroblock motion type is frame-

based, and the absolute value of motion vectors corresponding to the macroblock are less than or equal to the threshold.

15. The method of claim 7, in which the filtering is frame-based and operates on input samples from the top-field and bottom-field of the interlaced picture.

16. The method of claim 7, in which the filtering is field-based and operates on input samples from the top-field or bottom-field.

17. The method of claim 7, in which the filtering is field-based and operates on input samples from the bottom-field.

18. The method of claim 1, further comprising:  
encoding the progressive picture to an output video.

19. The method of claim 1, further comprising:  
rendering the progressive picture on a display device.

20. A system for processing a compressed input video, comprising:  
means for decoding the compressed input video to produce an interlaced picture, and macroblock coding information of the input video, the interlaced picture having a first spatial resolution, and a top-field and a bottom-field; and  
means for filtering, adaptively, the top-field and the bottom-field of the interlaced picture according to the macroblock coding information to produce a progressive picture with a second spatial resolution less than the first spatial resolution.

**ABSTRACT**

A method and system processes a compressed input video. The compressed input video is processed to produce an interlaced picture, and macroblock coding information of the input video. The interlaced picture has a first spatial resolution, and a top-field and a bottom-field. The top-field and the bottom-field of the interlaced picture are filtered adaptively according to the macroblock coding information to produce a progressive picture with a second spatial resolution less than the first spatial resolution.

**Representative Drawing**

Fig. 2

